Effects of rRNA Depletion

We have adapted the PacBio Iso-Seq protocol for use with prokaryotic samples by incorporating RNA polymerase-depletion and rRNA-depletion steps. In conjunction with SMRT Sequencing, which has average read-lengths of 10-18 kb, we are able to sequence full-length cDNA from individual transcripts and metatranscriptome samples.

In this method, we are able to detect full-length transcription of operons with alternative start and stop sites. In the area of metatranscriptomics, long reads reveal unambiguous gene sequences without the need for post-sequencing transcript assembly. Also described is a method for enriching primary transcripts in prokaryotes while at the same time removing rRNA more efficiently than current methods. This protocol can be used in conjunction with PacBio Iso-Seq in order to obtain full-length transcripts from prokaryotic transcripts and metatranscriptomes.

Sample Preparation Methods

**Figure 1.** (A) Representative traces of (A) total RNA, (B) polyA RNA and (C) polyA RNA-depleted RNA. PolyA RNA-depleted reaction was optimized in order to add ~200 nucleotides. PolyA RNA-depleted RNA showed good reduction in rRNA peaks and was the input for the cDNA synthesis reaction.

**Figure 2.** Sequence reads were mapped to the E. coli genome reference. Arrows show reduction in coverage of rRNA after rRNA depletion. Shaded peaks are most likely aberrant associated genes. Out of all traditional rRNA-depletion methods tested, Ribonuclease H had the highest rRNA-depletion efficiency.

**Figure 3.** Non-rRNA-depleted reads were mapped to the E. coli genome reference. Arrows show reduction in coverage of rRNA after rRNA depletion. Shaded peaks are most likely aberrant associated genes. Out of all traditional rRNA-depletion methods tested, Ribonuclease H had the highest rRNA-depletion efficiency.

**Figure 4.** Using long-read SMRT Sequencing, primary transcript start-stop sites can be identified. Full-length transcript reads re-map to (A) the x-axis and (B) the y-axis to E. coli. cDNA, showing multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

**Figure 5.** (A/B) Using long-read SMRT Sequencing, primary transcript start-stop sites can be identified. Full-length transcript reads re-map to (A) the x-axis and (B) the y-axis to E. coli. cDNA, showing multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

**Figure 6.** Enrichment for Primary Transcripts with NEB SMRT-Cappable-Seq. (A) Full-length transcripts, including transcripts with short starts, were able to be phasing the start and stop of the operon. (B) This method also allows phasing primary transcript start sites and allows for the detection of overlapping transcription start sites and polyadenylation sites.

**Figure 7.** Enrichment for Primary Transcripts with NEB SMRT-Cappable-Seq. (A) Full-length transcripts, including transcripts with short starts, were able to be phasing the start and stop of the operon. (B) This method also allows phasing primary transcript start sites and allows for the detection of overlapping transcription start sites and polyadenylation sites.

**Figure 8.** Enrichment for Primary Transcripts with NEB SMRT-Cappable-Seq. (A) Full-length transcripts, including transcripts with short starts, were able to be phasing the start and stop of the operon. (B) This method also allows phasing primary transcript start sites and allows for the detection of overlapping transcription start sites and polyadenylation sites.